J. Membrane Biol. 212, 109–117 (2006)
DOI: 10.1007/s00232-006-0870-z

The Journal of
Membrane
Biology

Genomic Organization of the Mammalian *SLC14a2* Urea Transporter Genes

C.P. Smith[1], **R.A. Fenton**[2]
[1]Faculty of Life Sciences, The University of Manchester, Oxford Road, Manchester, M13 9NT, UK
[2]The Water and Salt Research Center, Institute of Anatomy, Univeristy of Aarhus, Building 233/234 University of Aarhus, DK-8000 Aarhus, Denmark

Accepted: 17 April 2006

Abstract. Urea transporters encoded by the *UT-A* gene play fundamental roles in the kidney and possibly other tissues. Knowledge of the genomic organization of the mouse, rat and human *UT-A* genes has enabled the engineering of transgenic and knockout animals and these have helped refine our understanding of the role of UT-A proteins. This review summarizes the published work that has accrued on the structure and regulation of these genes. It also documents a novel cDNA, human UT-A3, which has enabled a major refinement of the human *UT-A* gene structure. This and other information contained in this review should prove useful for future comparative genomic analysis, studies addressing gene regulation and for the engineering of transgenic and knockout animal strains.

Key words: Membrane proteins — UT-A — Kidney — Genomic analysis

Introduction

Facilitative urea transporters regulate the passive movement of urea across cell membranes. In higher organisms these proteins are the products of two genes: *UT-A* (*SLC14a2*) and *UT-B* (*SLC14a1*). In fish, a third sub family, termed *UT-C*, has been described (Mistry et al., 2005). This review will focus on the mammalian *UT-A* genes. It will detail the genomic structures of the mammalian genes that have to date been defined, namely the mouse, rat and human *UT-A* genes. The gene structures presented in this review have been updated based on the most recent outputs from the respective genome projects and on recently discovered cDNAs.

Correspondence to: C.P. Smith; email: craig.smith@manchester.ac.uk

The Mammalian *UT-A* Genes

Over the past five years the genomic organization of the mouse, rat and human UT-A genes have been resolved (Bagnasco et al., 2001; Nakayama et al., 2001; Fenton et al., 2002b). The structures of the mouse and rat genes were assembled by conventional DNA-based methodologies, utilizing genomic library screening, prior to completion of the respective genomes. The human gene was assembled *in silico* based on two characterized UT-A cDNAs. Not surprisingly, the three genes share several characteristics. These include: their large size – all are greater than 400 kb; they reside on chromosome 18, in tandem with the *UT-B* gene; their transcription is driven by two promoters; they give rise to multiple mRNA isoforms as a result of differential promoter activity, alternative splicing and the use of alternate transcription and polyadenylation signals.

Although the genomic organization of the rat *UT-A* gene was described first, the structure of the mouse *UT-A* gene has proved fundamental to the development of UT-A knockout and transgenic animals; all of which have proved essential in understanding the physiological function of mammalian urea transporters. For these reasons, in this manuscript, the mouse gene will be described in detail first.

Organization of the Mouse *UT-A* Gene

In 1999, our group mapped the genomic location of the mouse *UT-A* gene by linkage analysis to chromosome 18 and found that it resided between genomic markers *D8Mit25* and *D8Mit186* (Fenton

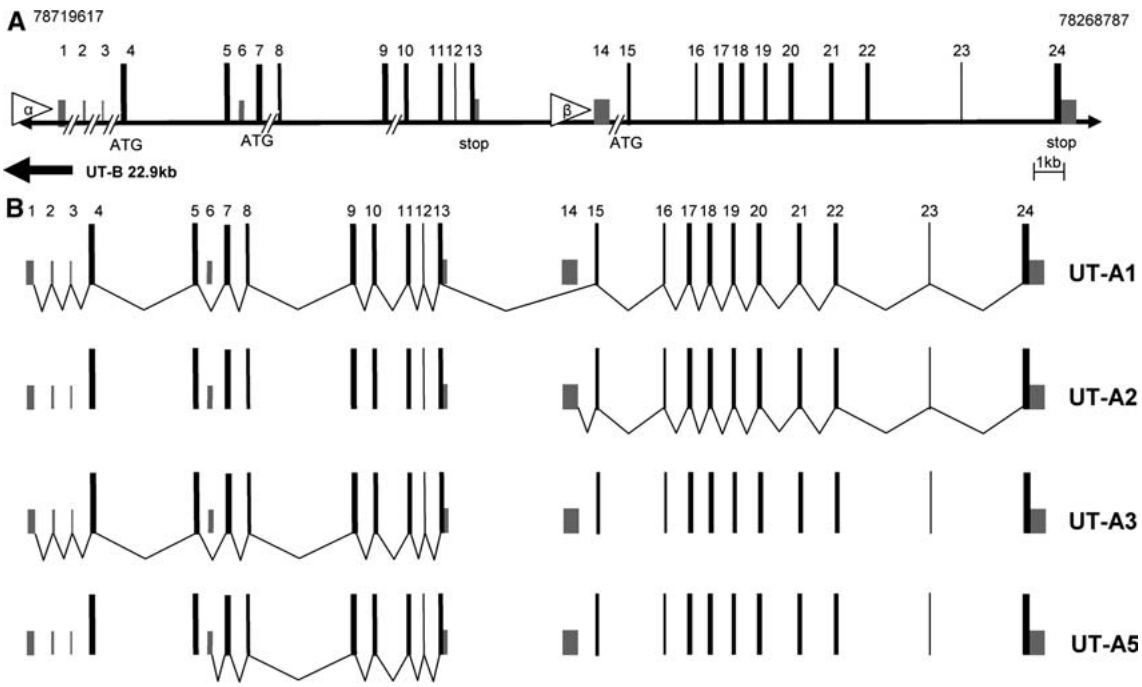

Fig. 1. The structure and gene products of the mouse *Slc14a2* gene. *A*. The gene comprises 24 exons interspersed over 450820 kb located on chromosome 18 at position q12.3. Coding exons (*black*) and untranslated exons (*grey*) are drawn to scale. Introns <5 kb are also drawn to scale. Introns >5 kb are denoted by //. Triangles denote the α and β promoters. The UT-B gene lies 22.9 kb from the 5′ of the gene. *B*. Splicing patterns are shown for the four characterized transcripts, UT-A1, UT-A2, UT-A3 and UT-5. The full coding sequence for mouse UT-A4 has not been resolved, therefore it is not included above. Coding exons (*black*) and untranslated exons (*grey*) are drawn to scale.

et al., 1999). Furthermore, we discovered that the *UT-A* gene occurred in tandem with the *UT-B* gene on chromosome 18 (Fenton et al., 1999). Resolution of the mouse genome has confirmed this location and has pin-pointed the gene to region q12.3 of chromosome 18, starting at base pair 78719617 and ending at base pair 78268787. The *UT-B* gene lies 22.9 kb proximal to the *UT-A* gene.

To determine the genomic organization of the mouse *UT-A* gene we initially screened a λFIXII/129sv mouse genomic library. We isolated a 16 kb genomic fragment that contained exons 9–13 (Fig. 1) and the UT-Aβ promoter (*see* below). Subsequently, three more lambda clones were isolated before it became apparent that the gene was greater in size than we had first anticipated. Thereafter, we changed our approach and screened a P1 artificial chromosome (PAC) genomic library. 'Shot-gun' sequencing of five isolated PACs yielded full coverage of the known UT-A coding sequences and many intronic regions. Based on this long and arduous process we determined that the gene consisted of 24 exons, had two promoters and spanned an estimated 300 kb (Fenton et al., 2002b).

The ensuing completion of the mouse genome project has enabled the initial genomic organization that we established to be validated and refined. Based on the mouse genome 'freeze' of March 2005 and using the mouse Blat software at http://genome.ucsc.edu/cgi-bin/hgBlat, the updated genomic organization of the mouse *UT-A* gene is presented in Fig. 1 and a revised table of intron sizes is included as Table 1. The mouse *UT-A* gene consists of 24 exons spread over 450820 bp. It is worthy to note that intron 2 and intron 3, which are 258 kb and 117 kb, respectively, account for more than 80% of the gene. At the 5′ end of the gene resides the UT-Aα promoter. This is followed by 13 exons and then the second promoter, the UT-Aβ promoter. Exons 14 to 24 lie 3′ to the UT-Aβ promoter.

To date, four mouse UT-A cDNAs have been isolated and characterized (Fenton et al., 2000, 2002a) and our knowledge of the exons encoded by the mouse *UT-A* gene is based on the sequence of these cDNAs. Figure 1 details the splicing pattern of the gene that gives rise to the known cDNAs. The mRNA encoding UT-A1, the largest UT-A protein (Fig. 2), is encoded by exons 1 to 5, 7 to 13, and 15 to 24. UT-A3 is encoded by exons 1 to 5 and exons 7 to 13. UT-A1 and UT-A3 are thought to share a common transcriptional start site upstream of exon 1 and a common translational start codon (AUG) located in exon 4. The translational stop codon for UT-A1 is found in exon 24, whereas that for UT-A3 is located in exon 13. Exon 13 is particularly interesting because it is a split exon. The first 175 bp of exon 13 contribute to the coding sequence of UT-A1; thereafter, exon 14 is spliced in. This has the effect of avoiding

Table 1. Revised organization of the mouse, rat and human UT-A gene *Slca14a2*

Mouse			Rat			Human		
Exon	Exon bp	Intron bp	Exon	Exon bp	Intron bp	Exon	Exon bp	Intron bp
1	240	11991	1	169	8451	1	201	12973
2	131	258192	2	230	238531	2	123	15086
3	66	117243	3	64	114637	3	106	16577
4	188	1843	4	184	2271	4	132	270042
5	181	453	5	180	995	5	89	131479
6	188	361	6	189	10539	6(1)	766?	UT-A1:9049 UT-A6:9085
7	190	9847	7	128	3674	7(2)	183	869
8	129	3442	8	192	6885	8(3)	180	1094
9	193	6204	9	147	1015	9(4)	189	5202
10	148	1058	10	134	390	10(5)	128	2343
11	135	373	11	49	463	11(5a)	127	2138
12	50	487	12	174(395/2382)§	2666	12(6)	192	2563
13	174(390)*	4327	13	707	16733	13(7)	147	1315
14	658	15256	14	125	3102	14(8)	134	1788
15	126	2256	15	49	677	15(9)	49	804
16	83	734	16	159	672	16(10)	175(422)†	19625
17	160	697	17	189	242	17(11)	21	2230
18	190	611	18	128	852	18(12)	83	716716
19	129	866	19	192	1561	19(13)	158	739
20	193	1523	20	147	612	20(14)	190	327
21	148	1374	21	134	3197	21(15)	127	829
22	135	3554	22	49	3709	22(16)	192	3402
23	50	3253	23	995	1405	23(17)	147	636
24	965		24	402		24(18)	134	5158
						25(19)	49	3295
						26(20)	787	

Exon and intron lengths are based on cDNA sequences. Exon and intron numbering is specific to each species and due to the complexity of the genes no attempt has been made to assign the same numbers to analogous exons or introns. Numbers in parentheses in the "Human Exon" column denote numbering for UT-A1.

* denotes mouse UT-A3 exon length in brackets; § denotes exon 12 lengths of rat UT-A3 and UT-A3b, respectively, in brackets; † denotes human UT-A3 exon length in brackets.

the stop codon (UAG) utilized by UT-A3 present at bp181–183 of exon 13. The use of a split exon to direct transcription of UT-A1 or UT-A3 is also evident in the rat and human *UT-A* genes.

The UT-A2 transcript is encoded by exons 14 to 24. This isoform has a unique transcriptional start site immediately 5′ of exon 14 and the proposed translation initiation codon (AUG) resides in exon 17. UT-A1 and UT-A2 share the same translational stop codon present in exon 24. Exon 14 is unique to UT-A2 and deletion of this exon and part of the UT-Aβ promoter was used to engineer UT-A2 knockout mice (Uchida et al., 2005) that are described elsewhere in this issue. The development of UT-A2 knockout mice has also provided some insight into the role of the UT-Aβ promoter. Deletion of this promoter does not affect the expression levels of UT-A1 or UT-A3, indicating that activity of the UT-Aβ promoter is not required for transcription of these splice variants and may therefore alone drive transcription of UT-A2.

It is worthy of note that a cDNA representing the full coding sequence for mouse UT-A4 has not been determined, although a cDNA fragment (Genbank accession no. AY221737) homologous to the central portion of rat UT-A4 has been isolated from caecum mRNA using reverse transcriptase PCR (RT-PCR), thus providing evidence that the transcript does indeed exist in mouse. However, since knowledge of the full-length cDNA or the corresponding protein product is lacking, it is not yet possible to determine the exon composition of mouse UT-A4.

UT-A5 is the newest member of the mouse UT-A family, and at 1.5 kb in length, is the shortest UT-A cDNA characterized to date (Fenton et al., 2000). It comprises exons 6 to 13 and is the only transcript known to utilize exon 6 (Fig. 1). It is this exon that contains the translational start AUG. Exons 1 to 5 are not incorporated in UT-A5, making the protein the equivalent of UT-A3 protein truncated by 139 amino acids at the amino-terminus.

Transcription of the Mouse *UT-A* Gene

Transcription of the mouse *UT-A* gene is driven by at least two promoters. A promoter in the 5′ flanking region, termed UT-Aα, controls the transcription of

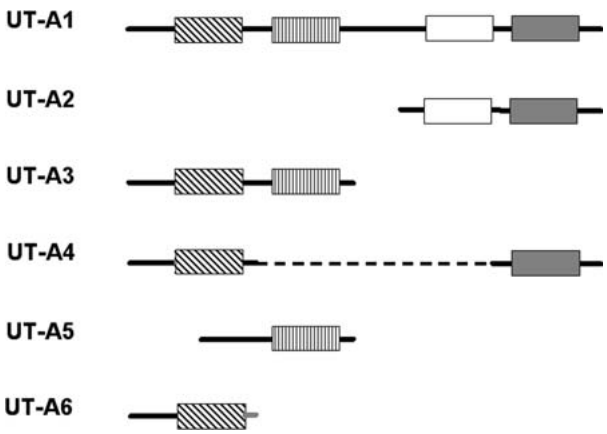

Fig. 2. The UT-A family. Schematic representation of the proposed primary structures of the UT-A proteins. *Continuous black lines* represent the primary sequence of each protein. UT-A1 is the largest UT-A protein. It comprises 931 amino acids (mouse) that confer four hydrophobic regions represented by shaded boxes. UT-A2 comprises the C-terminal 397 amino acids of UT-A1, UT-A3 is identical to the 461 amino acids of UT-A1. UT-A4 protein consists of the N-terminus and C-teminus of UT-A3 and UT-A2, respectively, and is proposed to be a 466-amino acid protein. UT-A5 is identical to UT-A3 except it is N-terminally truncated by 139 amino acids. The newest family member UT-A6 is identical to UT-A3 to amino acid 235, thereafter it contains a unique 19-amino acid cassette (*grey line*).

UT-A1, UT-A3 and possibly UT-A4. A second promoter, termed UT-Aβ, is situated in intron 14, upstream of the transcriptional start site of UT-A2. With respect to UT-A5, it appears possible that UT-A5 may be driven by a unique dedicated promoter that is potentially testis specific. The argument for this hypothesis rests on the fact that transgenic mice engineered to express β-galactosidase under the control of the UT-Aα promoter show β-galactosidase expression in only inner medullary collecting duct (IMCD) and the epithelia of the vas deferens (Fenton, Shodeinde & Knepper, 2005b), whereas no expression is detected in the testis where UT-A5 is known to be present. UT-A5 is expressed in the outer cells of the seminiferous tubules (Fenton et al., 2002a; Fenton et al., 2000) and since β-galactosidase was not detected in this region, it suggests that UT-A5 is driven by an altogether different promoter than UT-Aα. The caveat of this suggestion is that it is also possible that the promoter elements required for testis-specific expression were not part of the construct used to engineer the UT-Aα-β-galactosidase transgenic mice. Since computer analysis of the sequence immediately 5′ to exon 6 failed to identify motifs that are characteristic of mammalian promoters, empirical research is clearly required to determine if the sequence preceding this exon has the capacity to drive transcription.

The UT-Aα promoter resembles a TATA-less promoter, in that it lacks a TATA-box consensus sequence. However, it does contain three CCAAT boxes and consensus sequences for other well-characterized transcription factors, including two glucocorticoid response elements (GRE) and a tonicity-responsive enhancer/osmotic response element (TonE/ORE). Cloning of the UT-Aα promoter region into a luciferase reporter gene construct and subsequent expression in MDCK cells resulted in promoter activity that was increased by hyperosmolarity, mediated by the TonE domain, or cAMP (Fenton et al., 2002b). Interestingly, the response to cAMP is apparent in the absence of any cAMP response elements (CREs) consensi. In comparison, the UT-Aβ promoter contains a consensus TATA-box sequence and several other consensus sequences for well-characterized transcription factors, including five CRE motifs. Expression of a luciferase reporter gene construct containing the UT-A promoter region in MDCK cells resulted in low basal promoter activity. However, promoter activity was significantly increased by cAMP and this response was found to be CRE-mediated (Fenton et al., 2002b).

Knowledge of the mouse gene structure has enabled knockout mouse models to be engineered (*see* review in this issue). The first strain of mouse to be characterized was engineered by deleting exon 10 and the resulting mice lacked UT-A1, UT-A3 and UT-A5 (Fenton et al., 2004). Since UT-A1 and UT-A3 are the proteins responsible for transport of urea across IMCD epithelia, UT-A1/3/5 mice suffer from a urinary concentrating defect. This defect was pronounced in mice with high urinary urea excretion (due to feeding a high protein diet). In contrast, the mutant phenotype was significantly reduced when urinary urea excretion was low due to feeding a low protein diet (4%) (Fenton et al., 2005a). These observations were interpreted as evidence that urea acts as an osmotic diuretic if urea transport in the IMCD is absent (Fenton et al., 2004).

In addition to the above strain of mice, Uchida and colleagues have engineered a strain of mice that lack UT-A2 by deleting the UT-Aβ promoter and exon 14 (Uchida et al., 2005). UT-A2 is expressed in the thin descending limbs of the loop of Henle and was hypothesized to facilitate the recycling of urea within the renal medulla as part of the urinary concentrating mechanism (Sands, 2003). Despite its relatively high expression in the kidney and the proposed function, UT-A2 -deficient mice do not manifest a urinary concentrating defect even after 36 hours of thirsting (*see* review in this issue).

The most recent line of animals generated that utilizes the genomic information from the *UT-A* gene is a line of transgenic mice (termed UT-Aα-βGal mice) that express the reporter gene β-galactosidase (β-Gal) under the control of the mouse UT-Aα promoter (Fenton et al., 2005b). Within the kidney of UT-Aα-βGal mice, β-Gal expression was specific to the terminal portion of the papillary tip and co-localization

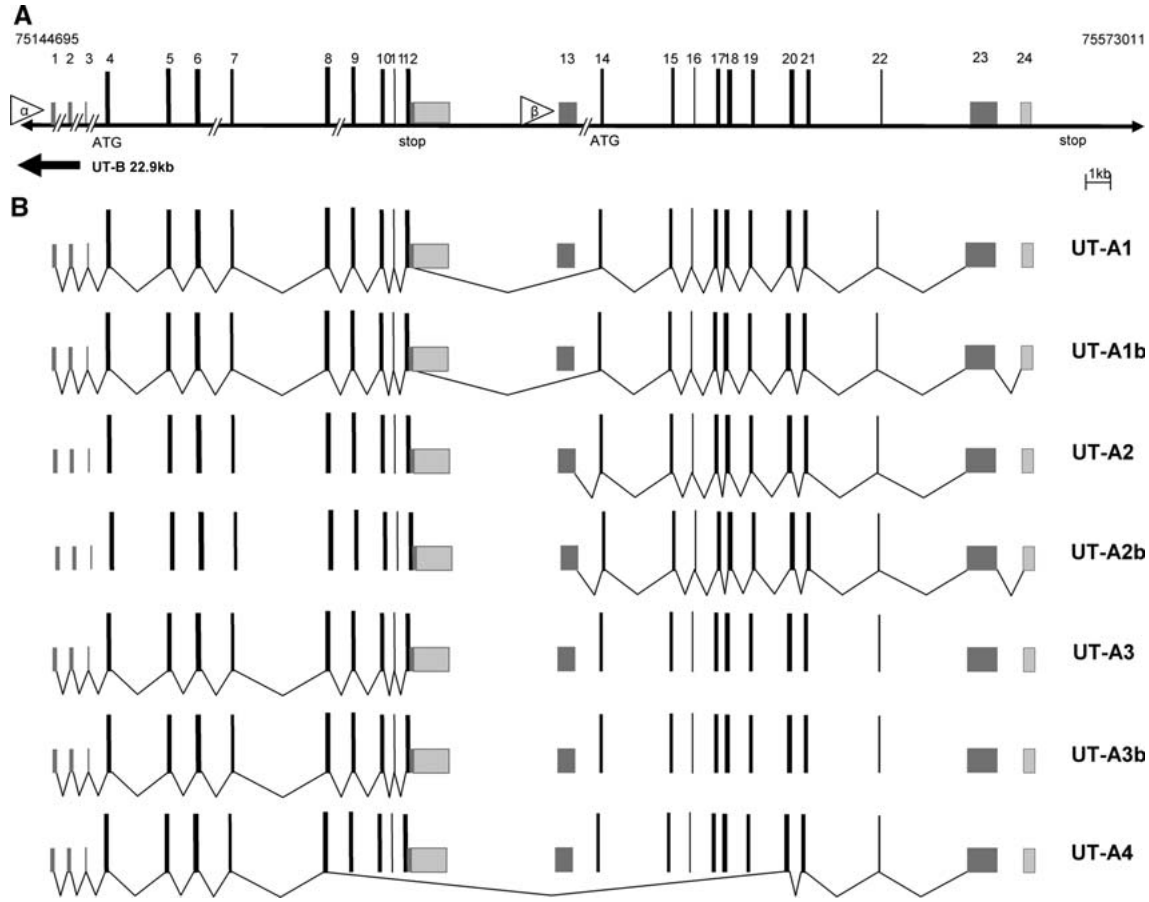

Fig. 3. The structure and gene products of the rat *Slc14aC* gene. (*A*) The gene comprises 24 exons spread over 428,319 kb of chromosome 18 at position q12.3. Coding exons (*black*) and untranslated exons (*grey*) are drawn to scale. Introns < 5 kb are also drawn to scale. Introns > 5 kb are denoted by //. Triangles denote the α and β promoters. The UT-B gene lies 22.9 kb away from the 5′ end of the gene. (*B*) Splicing patterns are shown for the seven characterized transcripts, UT-A1, UTA1b, UT-A2, UT-A2b, UT-A3, UT-A3b and UT-A4. Coding exons (*black*) and untranslated exons (*grey* and *grey hatched*) are drawn to scale. Exon 12 encodes UT-A1 amino acids and the 3′ UTR of UT-A3 (*grey*) and UT-A3b (*grey hatched*).

studies with AQP2 determined that expression was localized to the principal cells of the terminal IMCD. The pattern of β-Gal expression was consistent with the localization of UT-A1 and UT-A3 in the mouse kidney (Fenton et al., 2002c; Stewart et al., 2004). A novel finding from the characterization of the UT-Aα-βGal mouse line was that the UT-Aα promoter also targeted expression of β-Gal to the columnar epithelial principal cells within the vas deferens, a site not previously known to express urea transporters.

Several important questions concerning the mouse *UT-A* gene remain to be resolved. Firstly, does UT-A5 have a dedicated promoter region; secondly, do other transcripts exist that incorporate novel exons; finally, what is the effect of deleting all of the *UT-A* genes?

Organization of the Rat Gene

The genomic organization of the rat *UT-A* gene was the first to be resolved and was pieced together by Nakayama et al. (Nakayama et al., 2001) using conventional molecular biological methods. Five BAC genomic clones, four P1 genomic clones and one lambda phage genomic clone containing fragments of the rat gene were identified by hybridization to ^{32}P-labelled UT-A cDNAs. Intron-exon boundaries were determined using a combination of PCR and nucleotide sequencing. This strategy identified 24 exons that were estimated to be spread over more than 300 kb. Subsequent examination of the rat gene using Blat analysis (genome freeze June 2003, http://genome.ucsc.edu/cgi-bin/hgBlat) for this review has enabled refinement of this original structure. Based on the sequences of the characterized rat UT-A cDNAs (Fig. 3), the gene consists of 24 exons spanning 428,319 bp. The gene lies on chromosome 18 at position q12.3 between base pair 75,144,695 and 75,573,011. The UT-B gene lies 22.9 kb proximal to the *UT-A* gene. A revised structure of the rat UT-A gene is shown in Fig. 3 and Table 1 lists the revised intronic sizes.

Rat UT-A exons range from 49 bp (exons 11, 15 and 22) to 995 bp (exon 23), whereas introns vary in size from 242 bp (intron 17) to 238,531 bp (intron 2). As in the mouse and human, the latter, along with intron 3 (114,637 bp), are striking in terms of their size.

The rat gene encodes seven transcripts (Fig. 3), potentially encoding four protein isoforms, UT-A1, UT-A2, UT-A3 and UT-A4. cDNAs encoding these transcripts have been isolated by cDNA library screens and RT-PCR experiments. UT-A1, UT-A2 and UT-A3 are the major renal isoforms and are well characterized (Smith et al., 1995; Shayakul, Steel & Hediger, 1996; Karakashian et al., 1999; Shayakul et al., 2001). Although a cDNA encoding UT-A4 has been characterized, a protein with the characteristics of UT-A4 has yet to be unequivocally identified in rat tissue. Despite this fact, and for convenience, the splicing pattern of UT-A4 is shown in Fig. 3 alongside that of the six other characterized cDNAs.

The patterns of splicing of UT-A1, UT-A2 and UT-A3 in rat are basically as described for the mouse gene. However, the cDNA representing rat UT-A5 has not been isolated. Thus, the exon corresponding to mouse exon 6 has not been identified in rat and consequently the numbering of exons differs between rat and mouse for exons numbered 6 and greater.

cDNAs representing transcripts UT-A1b, UT-A2b and UT-3b have been isolated and shown to differ from UT-A1, UT-A2 and UT-A3 transcripts by incorporation of additional 3′ untranslated sequences. Since this additional sequence is non-coding, the products of these transcripts do not differ from those encoded by UT-A1, UT-A2 and UT-A3 transcripts. Transcripts UT-A1b and UT-A2b differ from the UT-A1 and UT-A2 transcripts in that they incorporate exon 24 (402 bp), whereas the additional sequence found in UT-A3b arises from exon 12. Exon 12, like exon 13 of mouse, is a split exon. The first 174 bp encode part of UT-A1, while this and an additional 121 bp encode the 3′ terminal exon of UT-A3. In rat, however, this exon extends a further 1987 bp, making it 2382 bp in length. These extra 1987 bp, when added to UT-A3, make up the UT-A3b transcript. Whether these additional transcripts occur in other species is not known, although uncharacterized UT-A transcripts are known to exist in both human and mouse (Smith et al., 2004; Stewart et al., 2004). Furthermore, the roles of these transcripts are not known, although the levels of UT-A2b and UT-A3b have been shown to be modulated in response to thirsting, implicating a role in the urinary concentrating mechanism (Bagnasco et al., 2000).

Like the mouse *UT-A* gene, the rat *UT-A* gene has two promoters (Nakayama et al., 2001). The α promoter (referred to by Nakayama and colleagues as Promoter 1) is 5′ to exon 1, and the β promoter (referred to by Nakayama and colleagues as promoter 2) lies within the intron between exon 12 and exon 13. Modulation of these promoters regulates transcription of the seven known *UT-A* gene products (Fig. 3). It is likely, but untested, that activity of the α promoter alone is required for transcription of UT-A1, UT-A1b, UT-A3, UT-A3b and UT-A4 and in contrast, transcription of UT-A2 or UT-A2b is driven by the β promoter.

The rat UT-Aα promoter, like the mouse, is a TATA-less promoter, and contains two CCAAT motifs and a TonE tonicity enhancer motif in the 1.3 kb immediately 5′ to the transcriptional start site (Nakayama et al., 2000). The rat UT-Aα promoter, like the mouse orthologue, is sensitive to changes in tonicity and this may drive transcription when the renal medullary hypertonicity is increased, as occurs during urine concentration. Chronic thirsting induces an increase in UT-A2 and UT-A3 mRNA, but not in UT-A1 mRNA (Smith et al., 1995; Bagnasco et al., 2000). If during antidiuresis hypertonicity drives the UT-Aα promoter, the finding that UT-A1 mRNA levels do not increase after 3 days of thirsting suggests that UT-A3 is in some way preferentially transcribed. This may necessitate regulation of alternative splicing, or the half-life of UT-A3 mRNA to be greater than that of UT-A1. The 1.3 kb 5′ region flanking rat exon 1 does not contain consensus CRE motifs although it is currently untested whether the promoter responds to cAMP. Therefore, it is unknown whether the mouse and rat UT-Aα promoters have different responses to cAMP.

As in the mouse, the rat UT-Aβ contains a TATA motif and has CRE motifs, but unlike the mouse it also has a TonE motif (Nakayama et al., 1999). Constructs containing the rat UT-Aβ promoter when transfected into IMCD3 cells are sensitive to cAMP and this response has been shown to be mediated via the CRE motifs. In contrast, hyperosmolality has no effect on UT-Aβ promoter activity, indicating that the TonE motif is nonfunctional.

The rat UT-Aα promoter contains several glucocorticoid response elements (GREs) (Peng, Sands & Bagnasco, 2002). Physiologically, glucocorticoids increase fractional urea excretion and decrease urea permeability and UT-A1 protein abundance in the IMCD. Cell culture studies have shown that dexamethasone administration significantly decreases the activity of the rat UT-Aα promoter in LLC-PK(1)-GR101 cells and also decreases UT-A1 and UT-A3 mRNA expression. However, deletion of the GREs in the UT-Aα promoter region has no effect on sensitivity of the promoter to glucocorticoids. Thus, it is unlikely that these GREs regulate the UT-Aα promoter (Peng et al., 2002).

Organization of the Human Gene

Previously, the human *UT-A* gene was thought to differ considerably from the rodent and mouse gene

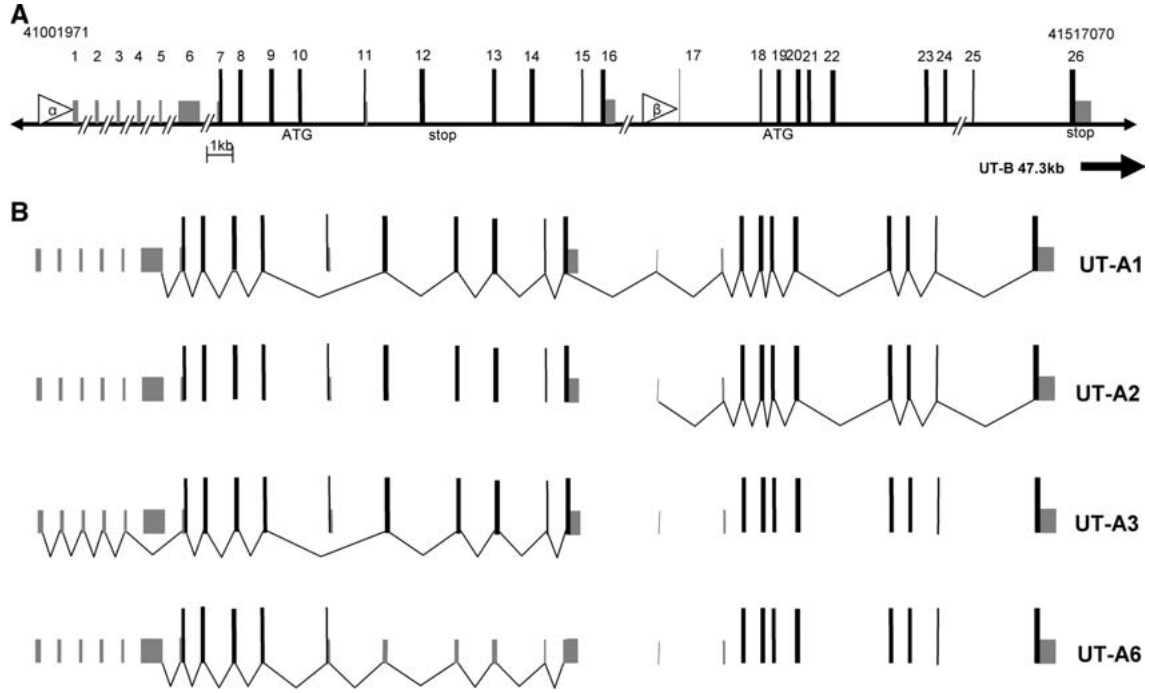

Fig. 4. Schematic representation of the genomic organization and gene products of the human urea transporter *UT-A* gene (*Slc14a2*). (*A*) The gene structure has been revised to incorporate exons belonging to the novel cDNAs encoding UT-A3 and UT-A6. The gene covers a total of 476,400 bp, from position 41,001,971 to 41,517,070 on chromosome 18 at position q12.3. The 26 exons are shown; exon width is representative of actual size and intronic distance is scaled. Introns > 5 kb are represented by // and are not scaled. All intron and exon sizes are listed in Table 1. *Triangles* denote the α and β promoters. The *UT-B* gene lies 47.3 kb to the 3′ end of the UT-A gene (*B*) Splicing patterns are shown of the human gene. Coding exons (*black*) and untranslated exons (*grey* and *grey hatched*) are drawn to scale. Exon 12 encodes UT-A1 amino acids and the 3′ UTR of UT-A3 (*grey*) and UT-A3b (*grey hatched*).

in that it was deemed to contain fewer exons and to cover only 67.5 kb (Bagnasco et al., 2001). This structure was based on exons from the two human cDNAs, *UT-A1*, and *UT-A2* that at the time were the only human isoforms to have been characterized. Since then, other human cDNAs have been identified and therefore a reassessment of the genomic organization was carried out for this review.

A novel cDNA, UT-A6 (Genbank accession no. AK074236), has been characterized from human colon (Smith et al., 2004) and a revised genomic structure was defined in the resulting publication to incorporate the extra exon that this cDNA contains (Fig. 4). Recent extensive searches of the available databases uncovered the sequence of an additional cDNA (Genbank accession no. AK122698). This novel 2271 bp cDNA, which originated from the New Energy and Industrial Technology Development's full-length human sequencing project (Institute of Medical Science, University of Tokyo, Japan), has 85% nucleotide identity to mouse and rat UT-A3. An open reading frame (ORF) starts at nucleotide 672 and extends to a UGA stop codon at 2025. This encodes a 451-amino acid protein with 82% amino acid identity to mouse and rat UT-A3 proteins. Based on this ORF, human UT-A3 is predicted to be 10 amino acids shorter than mouse UT-A3. This difference is also apparent when comparing mouse and human UT-A1, and thus is unlikely to be the result of sequencing errors. The ORF initiating sequence (AUG) employed by UT-A3 resides in exon 7. Exon 7 is also utilized by UT-A1 to initiate translation. However, exons upstream of this are not shared by UT-A1 and UT-A3. Exon 6 is, as far as we can tell, utilized only in UT-A1 and UT-A6, whereas exons 1,2,3,4 and 5 go to make up the 5′ untranslated region of the human UT-A3 transcript. It is these exons that, when incorporated into the genomic structure, increase the length of the gene from 67.5 kb to 476 kb. The existence of a human ortholog of UT-A3 has been suggested based on northern analysis and PCR (Bagnasco et al., 2001; Smith et al., 2004). A transcript estimated to be 2.4 kb was detected in human kidney medullary RNA (Bagnasco et al., 2001). The information presented in this review should aid future detailed analysis of this transcript and its corresponding protein.

The revised genomic organization of the human gene (Fig. 4) reveals that the gene is comprised of 26 exons spread over 476,400 bp. It resides on chromosome 18 at position q12.3, in tandem with the UT-B gene. The latter is 47.3 kb distal to the most 3′ human UT-A exon (exon 26). Like the rat and mouse genes, introns 2 and 3 are very large and account for

~85% of the gene. The next largest intron (19.6 kb), lies between exons 16 and 17. Upstream of this are exons encoding UT-A1, UT-A3 and UT-A6 and downstream of this are exons encoding UT-A1 and UT-A2. Within the intron flanked by exons 16 and 17 putatively lies the UT-Aβ promoter, although empirical verification of this is needed. Figure 4 shows the splicing patterns of the four known human UT-A gene products. As detailed earlier, UT-A1 and UT-A3 do not share the same exons for their 5′ UTR, but apart from exon 11, have in common exons 7 to 16. UT-A1 makes use of the first 175 bp of exon 16, then exons 19 to 26 are spliced in. Exon 16 is a split exon and is analogous to exon 12 in rat and exon 13 in mouse. Exon 26 contains the stop codon for UT-A1. In comparison, UT-A3 contains all of exon 16 and within the sequence that is not common with UT-A1 is the UT-A3 stop codon.

Interestingly, Bagnasco and colleagues were unable to find the location in the *UT-A* gene of the most 5′ 182 bp of human UT-A2 (Genbank accession no X96969). As part of the research carried out for this review, searches of the nucleotide databases also failed to assign a location to this DNA fragment despite completion of the human genome. Therefore, it remains a mystery as to the origin and genomic location of the most 5′ sequence of human UT-A2.

UT-A6 comprises exons 6 to 16 and, unlike UT-A1 or UT-A3, includes exon 11. Exon 11 contains an in-frame stop codon preceded by coding sequence for 19 amino acids that are unique, amongst the UT-A proteins, to UT-A6. Splicing of exon 11 causes termination of translation and yields a protein of 235 amino acids. The remaining exons (exons 12 to 16) are present in the UT-A6 transcript, but are non-coding and make up the 3′ UTR. At this point it should be noted that very little is known about the functional consequences of the alternative splicing of the *UT-A* gene. One can only assume that it serves to produce proteins targeted to specific membranes and/or with differing functional properties.

The human *UT-A* gene promoters have yet to be defined. However, it is possible that from the differences in the exons making up the 5′ ends of UT-A1 and UT-A3, these transcripts are driven by separate promoters. Experiments are required to determine if the sequences flanking exon 1 and exon 6 function as promoters. Based on observations made on the mouse and rat genes, transcription of UT-A2 is likely to be driven by the UT-Aβ promoter present in the intron proximal to exon 17. Exons 17 and 18 encode the 5′ UTR of the UT-A2 transcript. Exon 18 contains the translation initiation sequence AUG and thereafter the UT-A2 transcript includes exons 19 to 26. As for UT-A1, the UT-A2 stop codon resides in exon 26.

Concerted efforts by several research groups and the completion of multiple genome projects over the past five years have enabled us to gain a relatively good understanding of the genomic organization of the mouse, rat and human *UT-A* genes. This information has led to the engineering of UT-A transgenic and knockout mouse models that are helping us to refine our knowledge of both the renal and extrarenal roles urea transporters play. There is, however, still much to be done. Several important questions remain unanswered about the genes themselves. These include determining what drives transcription of UT-A5 and how is the splicing of the UT-A1/UT-A3 split exon controlled. In addition, mouse transgenic studies are underway to determine, amongst other things, the effect of deleting all of the *UT-A* gene and to engineer mice that transgenically express β-galactosidase driven by the UT-Aβ promoter. These and other experiments should keep researchers in this ever interesting field busy for some time to come.

References

Bagnasco, S.M., Peng, T., Janech, M.G., Karakashian, A., Sands, J.M. 2001. Cloning and characterization of the human urea transporter UT-A1 and mapping of the human Slc14a2 gene. *Am. J. Physiol.* **281:**F400–F406

Bagnasco, S.M., Peng, T., Nakayama, Y., Sands, J.M. 2000. Differential expression of individual UT-A urea transporter isoforms in rat kidney. *J. Am. Soc. Nephrol.* **11:**1980–1986

Fenton, R.A., Chou, C.L., Stewart, G.S., Smith, C.P., Knepper, M.A. 2004. Urinary concentrating defect in mice with selective deletion of phloretin-sensitive urea transporters in the renal collecting duct. *Proc. Natl. Acad. Sci. USA* **101:**7469–7474

Fenton, R.A., Cooper, G.J., Morris, I.D., Smith, C.P. 2002a. Coordinated expression of UT-A and UT-B urea transporters in rat testis. *Am. J. Physiol.* **282:**C1492–C1501

Fenton, R.A., Cottingham, C.A., Stewart, G.S., Howorth, A., Hewitt, J.A., Smith, C.P. 2002b. Structure and characterization of the mouse UT-A gene (Slc14a2). *Am. J. Physiol.* **282:**F630–F638

Fenton, R.A., Flynn, A., Shodeinde, A., Smith, C.P., Schnermann, J., Knepper, M.A. 2005a. Renal phenotype of UT-A urea transporter knockout mice. *J. Am. Soc. Nephrol.* **16:**1583–1592

Fenton, R.A., Hewitt, J.E., Howorth, A., Cottingham, C.A., Smith, C.P. 1999. The murine urea transporter genes Slc14a1 and Slc14a2 occur in tandem on chromosome 18. *Cytogenet. Cell Genet.* **87:**95–96

Fenton, R.A., Howorth, A., Cooper, G.J., Meccariello, R., Morris, I.D., Smith, C.P. 2000. Molecular characterization of a novel UT-A urea transporter isoform (UT-A5) in testis. *Am. J. Physiol.* **279:**C1425–C1431

Fenton, R.A., Shodeinde, A., Knepper, M.A. 2005b. The UT-A urea transporter promoter, UT-Aα, targets principal cells of the renal inner medullary collecting duct. *Am. J. Physiol.* **290:**F188–F195

Fenton, R.A., Stewart, G.S., Carpenter, B., Howorth, A., Potter, E.A., Cooper, G.J., Smith, C.P. 2002c. Characterization of mouse urea transporters UT-A1 and UT-A2. *Am. J. Physiol.* **283:**F817–F825

Karakashian, A., Timmer, R.T., Klein, J.D., Gunn, R.B., Sands, J.M., Bagnasco, S.M 1999. Cloning and characterization of two new isoforms of the rat kidney urea transporter: UT-A3 and UT-A4. *J. Am. Soc. Nephrol.* **10:**230–237

Mistry, A.C., Chen, G., Kato, A., Nag, K., Sands, J.M., Hirose, S 2005. A novel type of urea transporter, UT-C, is highly expressed in proximal tubule of seawater eel kidney. *Am. J. Physiol.* **288:**F455–F465

Nakayama, Y., Naruse, M., Karakashian, A., Peng, T., Sands, J.M., Bagnasco, S.M. 2001. Cloning of the rat Slc14a2 gene and genomic organization of the UT-A urea transporter. *Biochim. Biophys. Acta* **1518:**19–26

Nakayama, Y., Peng, T., Sands, J.M., Bagnasco, S.M. 1999. Both cAMP and tonicity increase the activity of the rat UT-A2 promoter. *J. Am. Soc. Nephrol.* **10:**21A

Nakayama, Y., Peng, T., Sands, J.M., Bagnasco, S.M. 2000. The TonE/TonEBP pathway mediates tonicity-responsive regulation of UT-A urea transporter expression. *J. Biol. Chem.* **275:**38275–38280

Peng, T., Sands, J.M., Bagnasco, S.M 2002. Glucocorticoids inhibit transcription and expression of the UT-A urea transporter gene. *Am. J. Physiol.* **282:**F853–F858

Sands, J.M.. 2003. Mammalian urea transporters. *Annu. Rev. Physiol.* **65:**543–566

Shayakul, C., Steel, A., Hediger, M.A. 1996. Molecular cloning and characterization of the vasopressin-regulated urea transporter of rat kidney collecting ducts. *J. Clin. Invest.* **98:**2580–2587

Shayakul, C., Tsukaguchi, H., Berger, U.V., Hediger, M.A. 2001. Molecular characterization of a novel urea transporter from kidney inner medullary collecting ducts. *Am. J. Physiol.* **280:**F487–F494

Smith, C.P., Lee, W.S., Martial, S., Knepper, M.A., You, G., Sands, J.M., Hediger, M.A. 1995. Cloning and regulation of expression of the rat kidney urea transporter (rUT2). *J. Clin. Invest.* **96:**1556–1563

Smith, C.P., Potter, E.A., Fenton, R.A., Stewart, G.S. 2004. Characterization of a human colonic cDNA encoding a structurally novel urea transporter, hUT-A6. *Am. J. Physiol.* **287:**C1087–C1093

Stewart, G.S., Fenton, R.A., Thevenod, F., Smith, C.P. 2004. Urea movement across mouse colonic plasma membranes is mediated by UT-A urea transporters. *Gastroenterology* **126:**765–773

Uchida, S., Sohara, E., Rai, T., Ikawa, M., Okabe, M., Sasaki, S. 2005. Impaired urea accumulation in the inner medulla of mice lacking the urea transporter UT-A2. *Mol. Cell Biol.* **25:**7357–7363